

Enhancing feature selection with a novel hybrid approach incorporating genetic algorithms and swarm intelligence techniques

Salsabila Benghazouani¹, Said Nouh¹, Abdelali Zakrani², Ihsane Haloum³, Mostafa Jebbar⁴

¹Department of Mathematics and Computer Science, Faculty of Sciences Ben M'Sick, Hassan II University, Casablanca, Morocco

²Department of Computer Science Engineering, ENSAM, Hassan II University, Casablanca, Morocco

³Department of Immunogenetics and Human Pathologies, Faculty of Medicine and Pharmacy, Hassan II University, Casablanca, Morocco

⁴Departement of Mathematics and Computer Science, EST, Hassan II University, Casablanca, Morocco

Article Info

Article history:

Received Oct 20, 2022

Revised Sep 2, 2023

Accepted Sep 15, 2023

Keywords:

Feature selection

Genetic algorithms

Machine learning

Multi-objective optimization

Swarm intelligence

ABSTRACT

Computing advances in data storage are leading to rapid growth in large-scale datasets. Using all features increases temporal/spatial complexity and negatively influences performance. Feature selection is a fundamental stage in data preprocessing, removing redundant and irrelevant features to minimize the number of features and enhance the performance of classification accuracy. Numerous optimization algorithms were employed to handle feature selection (FS) problems, and they outperform conventional FS techniques. However, there is no metaheuristic FS method that outperforms other optimization algorithms in many datasets. This motivated our study to incorporate the advantages of various optimization techniques to obtain a powerful technique that outperforms other methods in many datasets from different domains. In this article, a novel combined method GASI is developed using swarm intelligence (SI) based feature selection techniques and genetic algorithms (GA) that uses a multi-objective fitness function to seek the optimal subset of features. To assess the performance of the proposed approach, seven datasets have been collected from the UCI repository and exploited to test the newly established feature selection technique. The experimental results demonstrate that the suggested method GASI outperforms many powerful SI-based feature selection techniques studied. GASI obtains a better average fitness value and improves classification performance.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Salsabila Benghazouani

Department of Mathematics and Computer Science, Faculty of Sciences Ben M'Sick, Hassan II University
Casablanca, Morocco

Email: bengahazouani.salsabila239@gmail.com

1. INTRODUCTION

Feature selection plays a crucial role in the preprocessing phase of machine learning, it eliminates irrelevant and redundant features (noisy attributes), which increases the performance of a classifier and reduces the computational complexity [1]. In the healthcare sector, feature identification and selection play a vital role in enhancing accuracy in prediction, classification, and detection systems. This crucial preprocessing step not only enables reduction of dimensionality but also permits a better understanding of pathologies [2]. In an exhaustive search space, the number of possible combinations to determine the most relevant and non-redundant features is 2^n , where n represents the number of features (NP-complete problem)

[3]. Numerous feature selection algorithms have been suggested in the existing literature, and they are generally classified into three groups: filter algorithms, wrapper algorithms, and integrated algorithms [4]. The filter approach is independent on the learning algorithm and uses information-theoretic measures to assess and classify the features [5]. The advantage of this method is represented in terms of computation efficiency and it is power against overfitting [6]. By contrast, wrapper approaches employ a learning algorithm to assess subsets of features, which gives high accuracy for classifiers. However, they require a high computation time [7]. Integrated approaches integrate the advantages of filtering and wrapping methods. They incorporate the selection of variables during the learning process which allows to reach a compromise between the computational cost and the model performance [8]. However, the fundamental difficulty with the filter technique is that the features are chosen autonomously without using the machine learning classifier. While the wrapper technique chooses features using an optimization algorithm and works directly with the classifier [9]. Compared to the standard exhaustive search, optimization algorithms offer the advantage of efficiently selecting the optimal subset of features in a reasonable time.

Numerous optimization methods have recently been utilized to tackle feature selection (FS) problems, and they significantly outperform more traditional FS techniques. However, no meta-heuristic FS approach surpasses other optimization algorithms in many datasets. Such as, Rostami *et al.* [10] compared the performance of different swarm intelligence (SI) based feature selection methods on several datasets. The findings indicate that on the support vector machine (SVM) classifier, in the colon dataset, the cuckoo optimization algorithm (COA) outperforms the particle swarm optimization (PSO) method. However, in the isolated letter speech recognition (ISOLET) dataset, the PSO-based method performs better than COA. In study [11], an improved salp swarm algorithm (ISSA) is developed and compared to other swarm techniques. The findings revealed that when employing the k-nearest neighbor (KNN) classifier on the Waveform dataset, the SSA-based method exhibited superior performance in comparison to PSO. Conversely, for the Parkinson's dataset, the PSO-based method surpassed the SSA-based method in terms of performance.

This paper seeks to remedy these limitations by proposing a powerful feature selection approach based on a genetic algorithm (GA) that combines the advantages of various swarm intelligence (SI)-based feature selection techniques. The objective is to efficiently use helpful information from various SI-based feature selection techniques to obtain a better average fitness value and higher classification performance than other optimization algorithms in many data sets from different fields. The suggested feature selection approach has been applied to seven databases from the field of experimentation and publicly available UCI databases (colon, breast cancer Wisconsin, heart, arrhythmia, sonar, ionosphere, waveform). Moreover, the potency of the suggested method was then tested. The remainder of this article is structured to follow: section 2 describes the literature survey and related works. Section 3 details the proposed feature selection method GASI. Experimental results and discussion are shown in section 4, which is then succeeded by a conclusion and future perspectives in section 5.

2. LITERATURE SURVEY AND RELATED WORKS

A crucial issue for machine learning tasks is the large dimensionality of a data set with huge feature spaces and a limited number of samples [12]. Dimensionality reduction is a technique to tackle this issue by removing redundant and noisy features. This improves the classifier's performance and reduces its complexity in terms of computation and memory space. Dimensionality reduction approaches are typically categorized into two groups: feature selection and feature extraction. A reduction based on a data transformation is called a feature extraction, which replaces the initial data set with a new reduced one built from the initial set of features. A feature selection-based reduction chooses the most pertinent features from the dataset. In the following subsection, we briefly describe some feature selection approaches: filter, wrapper, and embedded.

Filter techniques use statistical performance measures to evaluate features and select the best ranking ones; these approaches are not dependent on the learning algorithm [5]. The filter methods are categorized into two categories: multivariate and univariate. The univariate approaches assess the pertinence of the attributes to the target class by an assessment criteria like mutual information (MI), information gain (IG), and Gini index (GI) [13]. This approach does not consider the interactions between features [3] and is prone to getting stuck in a local optimum [14]. The multivariate methods take into consideration the dependencies between features which allows the elimination of irrelevant and redundant variables. Among the multivariate methods, the maximum relevance minimum redundancy approach (MRMR) [15] and the relevance redundancy feature selection (RRFS) method [16].

The wrapper method is based on the learning algorithm to assess the variables to choose an optimum subset of characteristics with high classification precision [3]. Although this method uses a classifier and considers the interactions between variables, it remains computationally expensive [17]. Generally, a cross-validation mechanism is often used to reduce time complexity and avoid overfitting problems [18].

The embedded approach differs from other feature selection approaches. On the one hand, learning algorithms are not employed in the filter procedures. On the other hand, the wrapper approaches utilize a learning machine technique to assess the quality of feature subsets, independent of knowledge about the classification or regression function's specific structure [19]. Whereas the embedded approach integrates feature selection into the training process, it uses a machine learning method to seek the best subset of features while assuring a balance between computational cost and model performance [8]. Among the learning algorithms that use this concept: decision trees (DT), support vector machine (SVM), and AdaBoost. For example, DT is a tree-based classifier with several nodes and leaves. Each leaf is a label of class, while each node represents a particular feature. The relevance of a feature is determined by its location in the DT. Therefore, in DT-based integration approaches, the tree is first generated using an ensemble of models, and subsequently, the features engaged in the classification are chosen as the definitive subset of features [20].

Feature selection is considered among NP-hard problems due to the number of feasible subsets of variables increasing exponentially with the number of predictors [3]. Metaheuristics are approximate search methods often used for NP-hard problems, as they can achieve satisfactory (near-optimal) solutions in a short time [6]. Several feature selection methods use metaheuristics to escape local optimum and decrease computational complexity in high dimensional datasets [21]. This meta-heuristic is also based on an initially randomly generated population, then a fitness function that assesses the performance of the individual solutions of this one; a new population will be created if any termination criteria are not satisfied. This process is then iterated until one of the end criteria is fulfilled [22].

Genetic algorithm (GA) is an evolutionary computation algorithm that draws inspiration from the Darwinian evolution of biological populations. This well-known approach imitates the mechanism of natural selection, where the most appropriate individuals are selected for the reproduction of the next generation's children. This suggests that GA functions as child chromosomes are produced from their parents' chromosomes. Genetic operators, including crossover and mutation, are among the most crucial components of GAs and play a major part in utilizing the search space to discover novel solutions. While the mutation operators are in charge of creating new information by altering part of it, the crossover operators can search for new solutions using data already present in the population. In GAs, crossover operators are typically used to find novel solutions considerably more frequently than mutation operators. Although during the search procedure, the mutation operators assist in escaping the local optima. Genetic algorithms (GAs) have successfully shown their high ability to solve optimization problems, including feature selection problems [7], and numerous authors have suggested several GAs variants to solve the feature selection problem [4], [23]–[25]. In 2016, Cerrada *et al.*, [26] showed that GA could effectively achieve optimal global solutions for problems with large search spaces.

PSO is a robust optimization approach focused on SI, developed by Kennedy and Eberhart [27]. The approach is founded on the collective behavior of the shoal of fish and the flight of birds. PSO is utilized in various machine learning and feature selection applications. For instance, in [28], a multi-objective feature selection strategy using gray wolf optimization (GWO) and Newton's law derived PSO is created to reduce the number of chosen features and the rate of classification errors. In [29], a comparison of classification accuracy between PSO and a hybrid method that employs the Harris Hawk optimization algorithm (HHO) for optimizing SVM is performed. In [30], a hybrid meta-heuristic approach combining PSO and adaptive GA operators is introduced. This approach aims to optimize feature selection in machine learning models specifically designed to detect instances of tax avoidance. In [31], a thorough examination is conducted on current classification methods and gene selection techniques. The paper specifically emphasizes the effectiveness of emerging methods, like the SI algorithm, in the tasks of feature selection and classification for microarrays with high-dimensional data. In [32], a proposed system is presented that achieves automatic classification and detection of different pest attacks and plant infections. This is accomplished by employing a combination of radial basis probabilistic network (RBPN) and a genetic algorithm-based particle swarm optimization (GA-based PSO) method. In [33], a feature selection technique derived from PSO is proposed, which incorporates multiple classifiers. This approach utilizes adaptive parameters and strategies to tackle feature selection problems on a large scale, with the aim of improving classification accuracy and reducing computational complexity.

Differential evolution (DE) is a stochastic search algorithm focused on swarm intelligence, proposed by Storn and Price [34]. This optimization method was first developed to resolve the Chebyshev polynomial issue, but it also has been demonstrated effective in solving complex optimization issues [35]. Zhang *et al.* [36] presented a multi-objective feature selection method focused on differential evolution and defined a mutation operator to evade local optimum. Li *et al.* [37] proposed a novel large-scale multi-objective cooperative co-evolution method for feature selection to search for subsets of optimal features efficiently.

The cuckoo optimization algorithm (COA) is a recent evolutionary optimization approach focusing on swarm intelligence, which is derived from the life of a bird named the cuckoo [38]. Its principle is inspired by the behavior of the cuckoo bird in nesting and egg-laying to overcome optimization issues and find the global optimum [39]. In [40], a combination of neural network and cuckoo search algorithm is deployed for feature selection in heart disease classification. The firefly algorithm (FA) is an excellent instance of SI, in which underperforming entities collaborate to generate high-performance solutions. In [41], Yang introduced the FA with the basic notion being based on the optical connection between fireflies.

The salp swarm algorithm (SSA) is a recently developed algorithm based on SI that imitates the behavior of sea salps [42]. The SSA demonstrated a high performance when evaluated with various optimization issues. In [43], a new SSA and chaos theory combination is suggested to enhance feature selection accuracy. In [44], the dynamic salp swarm approach for feature selection is used to resolve the local optimum issue of SSA and to strike a balance between exploiting and exploring.

The Jaya algorithm (JA) is a recently implemented population-based meta-heuristic algorithm. Roa presented it in 2016 to handle constrained and unconstrained optimization problems. In [45], a novel hybrid feature selection approach is developed, incorporating the binary JA for the classification of microarray data is suggested to seek the optimum subset of features.

The flower pollination algorithm (FPA) is a meta-heuristic optimization technique that centers around the pollination process found in flowering plants. It was introduced by Yang in 2012 [46]. The primary goal of a flower is essentially to reproduce through the process of pollination, which involves the transfer of pollen and is frequently aided by pollinators such as birds and insects.

3. THE PROPOSED FEATURE SELECTION METHOD GASI

This section proposes a novel feature selection approach by integrating genetic algorithms and swarm intelligence-based feature selection techniques incorporating particle swarm optimization, differential evolution, cuckoo optimization algorithm, firefly algorithm, salp swarm algorithm, Jaya algorithm, flower pollination algorithm and other feature selection methods such as SelectFromModel and recursive feature elimination (RFE). The proposed approach GASI is founded on two principal pillars. The primary axis builds an initial smart population composed of the precious results of swarm intelligence-based feature selection algorithms (PSO, DE, COA, FA, SSA, JA, FPA, SelectFromModel, and RFE) that aim to discover the most optimal subset of features. The second axis introduces this intelligent population to the genetic algorithm in order to search for a better subset of features that contains a smaller number of features and improves the classification performance. The architecture of the suggested feature selection approach GASI is illustrated in Figure 1.

In this framework as shown in Figure 1, several SI-based feature selection techniques are applied to a dataset. Then an intelligent population composed of the feature subsets produced by these techniques is fed to the GA in the second step. The GA starts with this population and attempts to converge to the optimal subset of features employing genetic operators. An evaluation is made for each individual in the actual population based on a specified fitness function. A novel population is produced using genetic operations (selection, crossover, and mutation). This method is developed to maximize the classification accuracy and reduce the size of the feature subset. The following subsections describe the proposed GA method.

3.1. Encoding of individuals

In this context, individuals are represented using binary arrays consisting of n bits, where n corresponds to the number of features in the original dataset. A bit with a value of 1 in this array indicates the inclusion of the corresponding feature in the subset, while a bit with a value of 0 signifies the exclusion of that feature. This binary encoding method serves as an efficient means of representing feature subsets, enabling algorithms to make decisions about which features to include or exclude during various data analysis and optimization processes. It is a fundamental approach in feature selection and dimensionality reduction tasks.

3.2. Smart population with SI-based feature selection

Instead of creating an initial population with a predetermined number of randomly generated individuals, we take advantage of the best solutions obtained by many powerful SI-based feature selection approaches. For that purpose, an intelligent population is constructed from the optimal subsets of features produced by the different SI-based feature selection approaches, with additional randomly generated individuals to keep the diversity of the next generation. This intelligent collection of features will be fed into the genetic algorithm as the initial population to search for the optimal subset of features that maximizes classification performance and reduces the size of the features.

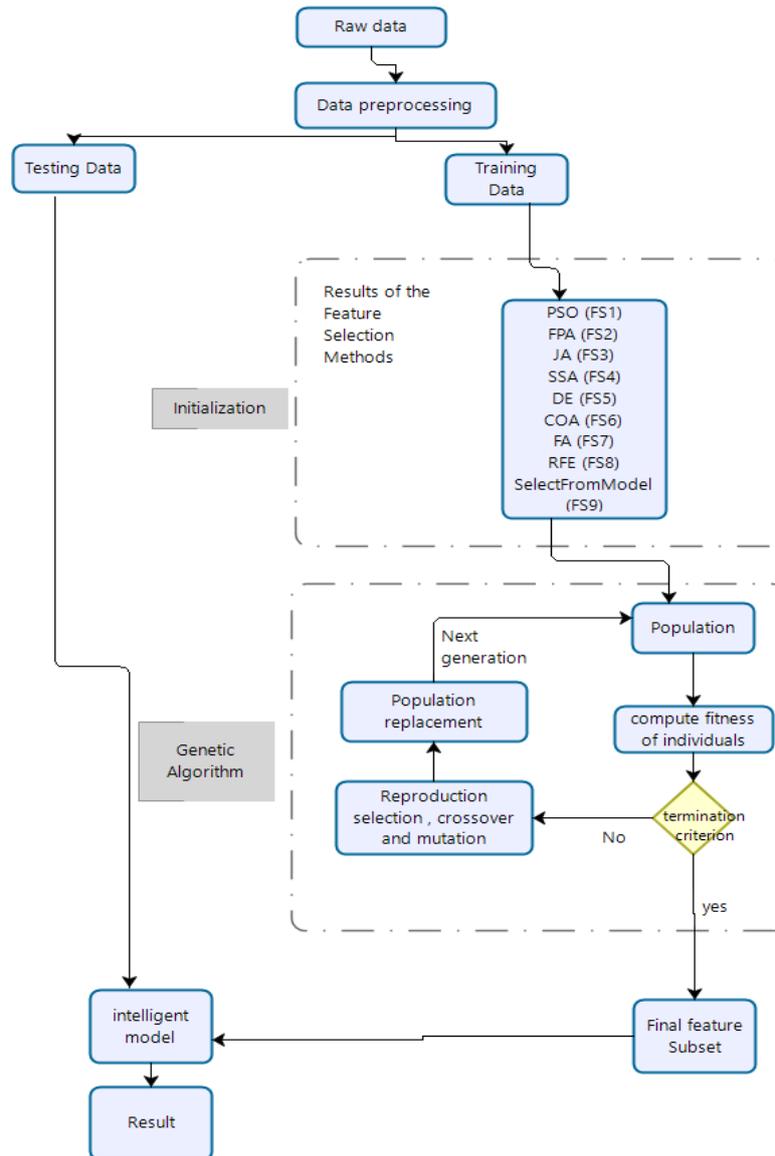


Figure 1. Flowchart of the proposed method GASI

3.3. Fitness function

The fitness function operates by simultaneously considering two distinct objectives: the enhancement of classification accuracy and the reduction of the number of selected features. To convert this function into a minimization problem, we introduce weights for each of these objectives. These weights enable the amalgamation of these criteria into a unified representation of the fitness function. Consequently, the fitness function can be articulated in the following manner:

$$\text{Min } F(X) = \alpha \cdot Er_c(X) + \beta \cdot P_{FS}(X) \tag{1}$$

In (1) represents the fitness function that assesses the fitness value attached to each individual. Where X is a vector of features illustrating a selected subset of features, α and β in (1) are the weights assigned to each objective, the classification error, and the proportion of features selected, respectively, which fulfill the below conditions in (2) and (3).

$$\beta = 1 - \alpha \tag{2}$$

$$\alpha \in [0,1] \tag{3}$$

The parameters α and β , where $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$, are utilized to regulate the importance of classification accuracy and feature reduction. The values of α and β used in previous studies [47], [48] are also employed in the current experiments, with α set to 0.99. In (4) defines the error rate of classification $Er_c(x)$ that needs to be minimized and is complementary to the accuracy of the classifier defined in (5), while $P_{FS}(x)$ represents the proportion of selected predictors in (6), where D is the size of the individual and f_i is a binary variable that specifies whether the feature I is present or not in a selected individual (7).

$$Er_c = 1 - Accuracy \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$0 \leq P_{FS} = \frac{\sum_{i=1}^D f_i}{D} \leq 1 \tag{6}$$

$$f_i \in 0,1 \tag{7}$$

where TP, TN, FP, and FN correspond to the number of true positives, true negatives, false positives, and false negatives, respectively. The individuals are ranked according to their fitness value, and the chosen number of individuals with the lowest fitness value are considered the parents of the next generation.

3.4. Genetic operators

3.4.1. Selection

Once a population of chromosomes has been created, GA will search for a few pairs of parent chromosomes to apply a crossover operation. To select the parent chromosomes, we employed the Roulette selection approach [49]. Every chromosome gets space and place on the roulette wheel based on its fitness. The wheel is rotated after the chromosomes have been placed on it. When the wheel stops turning, a random pointer on it points at the chosen chromosome. The best chromosomes have a large space, implying a high probability of being selected. The likelihood of selecting individual x is proportional to its fitness and determined by (8):

$$P_{Sel}(x_i) = \frac{fitness(x_i)}{\sum_{j=1}^N fitness(x_j)} \tag{8}$$

where $fitness(x_i)$ is the fitness value of the individual x_i . An elitist approach guarantees that the better individual is systematically moved to the next generation with no crossover or mutation. It is essential to maintain the constant convergence of the genetic algorithm. Tournament selection where individuals are randomly selected and compete for survival allows for the inclusion of individuals with lower fitness values, promoting diversity by giving them a chance to contribute to the next generation.

3.4.2. Crossover

Crossover is applied to each pair of chromosomes selected by the abovementioned method with a specified probability P_c . A high probability P_c involves the appearance of new individuals in the population. The crossover is applied by selecting a random point on the chromosome where the exchange of the parent's parts occurs. This process then gives rise to a new offspring based on the selected exchange point with particular parts of the parents as shown in Figure 2.

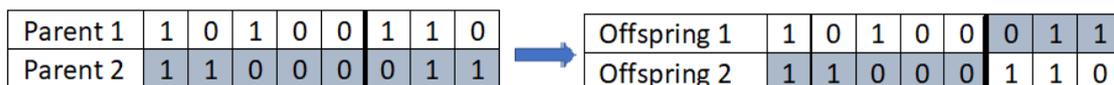


Figure 2. One point crossover

3.4.3. Mutation

The term “mutation” refers to the random change in the value of a gene on a chromosome. Mutation acts as a background noise that prevents evolution from freezing. It extends space exploration and ensures that the global optimum can be reached. Therefore, this operator avoids converging to local optima. The technique used is a uniform mutation, so each bit of a chromosome has a low probability P_m of being flipped.

4. RESULTS AND DISCUSSION

This section aims to assess the proposed genetic algorithm in terms of fitness values, feature space reduction, and prediction accuracy. The proposed method GASI uses an initial population based on all the feature selection techniques chosen in this paper (FPA, JA, SSA, DE, COA, FA, PSO, RFE, and SelectFromModel). We have also compared these results with other SI-based feature selection techniques, which would allow a robust empirical study. It is worth noting that even if this genetic algorithm's computational cost is higher than other feature selection approaches, they surpass them in terms of fitness value and accuracy. The rest of this section describes the employed datasets, the used classifiers, the evaluated approaches, the evaluation measures, the results, and the discussion in the following subsections.

4.1. Datasets

Based on the World Health Organization (WHO), heart disorders and cancer are the two leading causes of mortality in developing and under-developed countries. Breast, lung, colon, and rectum tumors remain the most commonly diagnosed cancers worldwide [50]. For these reasons, we tested the proposed method on seven well-known datasets to evaluate its performance. These datasets are colon, breast cancer Wisconsin, heart, arrhythmia, sonar, ionosphere, and waveform collected from the UCI repository. The description of the datasets is given in Table 1.

Table 1. Description of the seven studied datasets

Dataset	No. of features	No. of instances	No. of classes
Colon cancer	2,000	62	2
Breast cancer Wisconsin	32	569	2
Heart	13	270	2
Arrhythmia	279	452	16
Sonar	60	208	2
Ionosphere	34	351	2
Waveform	21	5,000	3

4.2. Classifier description

As learning algorithms, we used two popular classifiers, namely logistic regression (LR) and AdaBoost (AB), to evaluate the proposed method. LR is a method for predicting a dichotomous dependent variable. This approach finds the best fitting model that describes the association between the attributes of the dependent variable and a set of independent variables [51]. AdaBoost, an abbreviation for adaptive boosting, is a meta-algorithm for machine learning proposed by Freund and Schapire [52]. A classifier AdaBoost is a meta estimator that first fits a classifier and adapts it in multiple instances on the same dataset. Subsequently, the weights of the misclassified samples are adjusted to prioritize severe cases, leading to subsequent classifiers focusing more on them.

4.3. The evaluated methods

In the experiments, we utilized SI-based feature selection approaches, specifically FPA, JA, SSA, DE, COA, FA, and PSO defined in section 2. Additionally, we incorporated two more selection methods, SelectFromModel and recursive feature elimination (RFE), which will be described subsequently. This diverse set of feature selection methods was chosen to comprehensively explore their effectiveness within the experimental context, enabling a thorough examination of their impact on the overall results. The inclusion of these various methods provides a robust foundation for assessing the role of feature selection in the study's outcomes.

SelectFromModel is one of the feature selection techniques for extracting essential and relevant features. It removes features whose corresponding importance values are below the given threshold value. This model works with estimators that have important features or coefficients [53].

RFE is an integrated technique compatible with various learning algorithms like SVMs and Lasso. Its primary function involves iteratively and explicitly reducing the number of features by recursively eliminating those with low weights or importance scores. RFE is particularly useful for optimizing model performance by focusing on the most relevant features in a dataset [54].

Table 2 displays the average accuracy of the classification (in%), over ten trials, of SelectFromModel and RFE using the logistic regression and AdaBoost classifiers. The results indicate that in colon cancer, breast cancer, and waveform datasets, the SelectFromModel method slightly improves the classification performance compared to the RFE method. In contrast, the latter enhances the classification accuracy in the remaining datasets.

Table 2. A comparison between SelectFromModel and RFE in terms of classification accuracy using the logistic regression and AdaBoost classifiers

Dataset	Classifier	Feature selection methods	
		SelectFromModel	RFE
Colon cancer	logistic regression	85.78	85.78
	AdaBoost	86.84	85.78
Breast cancer	logistic regression	96.30	95.90
	AdaBoost	95.70	95.36
Heart	logistic regression	83.95	86.41
	AdaBoost	66.66	74.07
Arrhythmia	logistic regression	65.44	66.91
	AdaBoost	63.97	63.97
Sonar	logistic regression	87.30	84.12
	AdaBoost	87.30	88.88
Ionosphere	logistic regression	84.90	93.39
	AdaBoost	93.39	89.62
Waveform	logistic regression	85.79	85.06
	AdaBoost	83.26	82.26

4.4. Proposed method GASI parameters

As the parameters significantly impact the efficiency of the genetic algorithms, they should be chosen carefully to obtain the highest performance. Table 3 presents the parameters employed in GASI evaluation. The mentioned values were determined empirically through several experiments of the proposed approach.

Table 3. Common parameters for the proposed method

Parameter	Value
Population size	30
Number of generations	100
P_c (crossover probability)	0.8
P_m (mutation probability)	0.01
Weight of the classification error α	0.99
Weight of the number of selected predictors β	0.01
Elitist strategy	The best individual goes to the next iteration
NR_{max} : the number of runs maximal	10

4.4.1. Evaluation measures

The effectiveness of the proposed strategy GASI was evaluated in terms of the average accuracy of the classifier, the minimum number and rate of remaining features as in (12)-(14), the average, the best, and the worst fitness values as in (9)-(11). The proposed strategy is then compared with other metaheuristic algorithms using these measures. A mathematical formula for the evaluation measures is given in (9) to (14).

Average fitness value

$$Avg_f = \frac{1}{NR_{max}} \sum_{j=1}^{NR_{max}} f_j^* \quad (9)$$

Optimal value for fitness

$$Best_f = \min \{ f_j^* : j = 1, 2, \dots, NR_{max} \} \quad (10)$$

Worst value for fitness

$$Max_f = \max \{ f_j^* : j = 1, 2, \dots, NR_{max} \} \quad (11)$$

Average accuracy

$$Avg_{Acc} = \frac{1}{NR_{max}} \sum_{j=1}^{NR_{max}} Acc_j^* \quad (12)$$

Average number of selected features

$$Avg_{NFS} = \frac{1}{NR_{max}} \sum_{j=1}^{NR_{max}} length(x)_j^* \quad (13)$$

Rate of the remaining features

$$Remain_{FR} = \frac{1}{NR_{max}} \sum_{j=1}^{NR_{max}} \frac{length(x)_j}{TN} \quad (14)$$

where NR_{max} is the maximum number of runs and f_j^* represents the best fitness score attained at the j^{th} run. Acc_j^* is the optimum accuracy of the classifier achieved at the j^{th} run, $length(x)_j$ indicates the number of features that have been selected, and TN denotes the total number of features in the given dataset.

4.5. Experimental evaluation

In this sub-section, the performances of the suggested method are evaluated and compared to other powerful competitors. The results are presented in terms of fitness values, remaining feature rates, and classification accuracy. Each feature selection approach is executed ten times in each experiment, and the average of these different runs is utilized for comparing the different approaches. In addition, each dataset is normalized and randomly divided into a training set (70% of the dataset) and a testing set (30%). All these approaches are executed using Python on an Intel Core- i7 CPU with 16 GB of RAM.

Table 4 displays the average classification accuracy and rate of the remaining features (in %) over ten runs of the suggested approach GASI and the different SI-based feature selection approaches (i.e., FPA, JA, SSA, DE, COA, FA, and PSO) using the LR and AB classifiers, the best results are indicated in bold. The results in Table 4 show that the suggested approach is more optimal than many other SI-based feature selection techniques. It was able to select fewer features in most datasets while increasing the classification performance. Table 4 demonstrates that the proposed method GASI consistently outperforms the other SI-based feature selection technique. For instance, in the Colon cancer dataset using the RL classifier, GASI method achieved a classification accuracy of 98.94%. Contrarily, these values were reported as 94.37, 96.31, 97.89, 94.73, 95.25, 94.73, and 95.25, respectively, for the FPA, JA, SSA, DE, COA, FA, and PSO approaches. In addition, the AdaBoost classifier enhanced the classification accuracy to 100% for the suggested approach GASI. However, the accuracy of the FPA, JA, SSA, DE, COA, FA, and PSO methods was 92.62, 90.52, 95.25, 93.15, 99.47, 95.25, and 95.26, respectively. Table 4 also presents the number of selected features. The results show that all methods significantly reduce the dimensionality by only selecting a small part of the original features. For instance, by employing the RL classifier, the GASI approach performs better than the other SI-based methods in the colon cancer and sonar datasets, selecting only 0.2746 and 0.2694, respectively. Furthermore, the PSO method chose an average of 0.2343 and 0.3788 features in the breast cancer and arrhythmia datasets, respectively. However, the FPA method chose an average of 0.3615 features in the heart dataset, compared to the JA method's average of 0.2411 features in the Ionosphere dataset and the DE method's average of 0.7190 features in the waveform dataset.

Table 5 presents the evaluated results in terms of the average, the best (minimum), and the worst (maximum) fitness values. The results reveal that the proposed method, GASI, performed better in all datasets than other SI-based feature selection algorithms and delivered the smallest average fitness function value. For example, in the colon cancer dataset employing the RL classifier, the GASI approach provided an average value of 0.0131 in the fitness function. On the other hand, for the FPA, JA, SSA, DE, COA, FA, and PSO methods, these values were 0.0569, 0.0249, 0.0255, 0.0560, 0.0512, 0.0568, and 0.0507, respectively. Using the AdaBoost classifier, the mean value of the fitness function is 0.0037 for the GASI method. However, the values of other methods FPA, JA, SSA, DE, COA, FA and PSO were 0.0777, 0.0253, 0.0464, 0.0724, 0.0099, 0.0516 and 0.0044 respectively.

Figures 3 and 4 illustrate the mean classification accuracy on all datasets for the RL and AdaBoost classifiers, respectively. From these results, we can observe that, on all classifiers, the proposed method GASI obtained the highest average classification accuracy. The results in Figure 3 show that the GASI method achieved an average classification accuracy of 92.52% which ranked first with a margin of 2.89% compared to the JA approach, which achieved the second-best average classification accuracy. The FA method scored third with a margin of 3.54% compared to the best method. Furthermore, according to the results in Figure 4, on the AB classifier, the suggested approach GASI obtained the first place with an average classification accuracy of 91.53% with a margin of 1.81% compared to the COA method, which achieved the second-best average classification accuracy. In contrast, the DE approach secured the third position with an average classification accuracy of 89.12%.

Figure 5 provides an in-depth analysis of average fitness values across various datasets, with Figures 5(a) to 5(g) offering specific comparisons of mean fitness values for datasets such as colon cancer, breast cancer, heart, sonar, ionosphere, waveform, and arrhythmia. The results consistently affirm the superior performance of our proposed GASI approach when compared to other SI-based feature selection methods. This superiority is consistently demonstrated by GASI achieving the smallest average fitness value during evaluations conducted using both logistic regression and AdaBoost classifiers. These findings accentuate GASI's effectiveness in enhancing feature selection and its potential for wide-ranging applications across diverse datasets and machine learning algorithms.

Table 4. Average classification accuracy and remaining feature rates of the different feature selection approaches with logistic regression and AdaBoost classifier. The best results are marked in bold

Dataset	N_features	Method	Classifier					
			Logistic regression			AdaBoost		
			Avg _{Acc} (%)	Avg _{NSF}	Remain _{FR}	Avg _{Acc} (%)	Avg _{NSF}	Remain _{FR}
Colon cancer	2,000	GASI	98.94	549.2	0.2746	100	750.8	0.3754
		FPA	94.37	961.1	0.4805	92.62	966.6	0.4833
		JA	96.31	815.3	0.4076	90.52	911.5	0.4557
		SSA	97.89	950.9	0.4754	95.25	958	0.479
		DE	94.73	779.6	0.3898	93.15	938.7	0.4693
		COA	95.25	864.3	0.4321	99.47	944.8	0.4724
		FA	94.73	941.3	0.4706	95.25	955.8	0.4779
		PSO	95.25	765	0.3825	95.26	887.2	0.4436
		GASI	99.79	11.6	0.3625	100	8.9	0.2781
Breast cancer	32	FPA	96.23	12.1	0.3781	98.98	14.7	0.4593
		JA	97.17	7.8	0.2437	97.24	11.9	0.3718
		SSA	95.90	13.6	0.4468	97.37	14	0.4375
		DE	97.24	11.7	0.3656	99.45	11.9	0.3718
		COA	96.70	10.5	0.3281	98.18	11	0.3437
		FA	98.11	11.5	0.3593	99.32	11.6	0.3625
		PSO	96.50	7.5	0.2343	97.24	12	0.375
		GASI	91.35	7	0.5384	91.35	10	0.7692
		FPA	83.95	4.7	0.3615	89.13	8.1	0.6230
Heart	13	JA	90.37	9.5	0.7307	89.87	5.7	0.4384
		SSA	89.01	5.6	0.4307	87.77	6.3	0.4846
		DE	89.13	8.7	0.6692	89.25	5.6	0.4307
		COA	90.12	5.3	0.4076	89.13	5.4	0.4153
		FA	91.11	5.2	0.4	87.65	3	0.2307
		PSO	85.18	6.4	0.4923	89.38	7.3	0.5615
		GASI	78.67	130	0.4666	69.11	105.3	0.3774
		FPA	68.67	130.4	0.4673	66.02	129.5	0.4641
		JA	69.63	110.3	0.3953	64.63	72.7	0.2605
Arrhythmia	279	SSA	68.60	134.8	0.4831	63.89	121.5	0.4354
		DE	70.00	109.4	0.3921	65.44	76.7	0.2749
		COA	69.41	123.3	0.4419	66.98	101.5	0.3637
		FA	69.11	136.1	0.4878	67.86	129.9	0.4655
		PSO	69.77	105.7	0.3788	66.17	79.9	0.2863
		GASI	93.96	15.9	0.2694	98.41	28	0.4745
		FPA	90.63	28.3	0.4796	93.17	29.5	0.5
		JA	92.85	19.3	0.3271	93.65	26.3	0.4457
		SSA	92.69	25.9	0.4389	91.58	28	0.4745
Sonar	60	DE	89.52	19.7	0.3338	96.66	26.8	0.4542
		COA	88.25	25.2	0.4271	95.39	28.1	0.4762
		FA	89.52	26.2	0.4440	94.70	28.88	0.4406
		PSO	92.38	22.7	0.3847	94.44	25.1	0.4254
		GASI	96.22	11	0.3235	96.22	13.5	0.3979
		FPA	93.11	15.9	0.4676	93.01	14.2	0.4176
		JA	93.58	8.2	0.2411	94.81	11.8	0.3470
		SSA	90.37	15.9	0.4676	95.37	14.6	0.4294
		DE	91.50	11.5	0.3382	94.90	12.1	0.3558
Ionosphere	34	COA	93.67	13.9	0.4088	94.15	13.7	0.4029
		FA	93.77	12.6	0.3705	93.86	13.1	0.3852
		PSO	93.30	13	0.3823	94.71	12.4	0.3647
		GASI	88.73	17	0.8095	85.64	16.5	0.7857
		FPA	86.07	15.6	0.7428	84.39	15.5	0.7380
		JA	87.53	17	0.8095	85.20	15.3	0.7285
		SSA	86.22	15.8	0.7523	84.39	14.9	0.7095
		DE	87.05	15.1	0.7190	84.96	14.8	0.7047
		COA	87.27	16.6	0.7904	84.72	14.9	0.7095
Waveform	21	FA	86.51	17.3	0.8238	85.11	15.6	0.7428
		PSO	87.03	16.2	0.7714	85.46	15.2	0.7238

Table 5. Average, best, and worst fitness values of the different feature selection methods using logistic regression and AdaBoost classifier. The best results of fitness values are indicated in bold

Dataset	Method	Classifier					
		Logistic regression			AdaBoost		
		Avg _f	Best _f	Max _f	Avg _f	Best _f	Max _f
Colon cancer	GASI	0.0131	0.0009	0.0529	0.0037	0.0032	0.0040
	FPA	0.0569	0.0566	0.0570	0.0777	0.0567	0.1089
	JA	0.0249	0.0039	0.0562	0.0253	0.0044	0.0566
	SSA	0.0255	0.0047	0.0568	0.0464	0.0048	0.0569
	DE	0.0560	0.0558	0.0564	0.0724	0.0047	0.1610
	COA	0.0512	0.0046	0.0564	0.0099	0.0046	0.0567
	FA	0.0568	0.0567	0.0569	0.0516	0.0047	0.1088
Breast cancer	PSO	0.0507	0.0037	0.0560	0.0044	0.0034	0.0047
	GASI	0.0058	0.0033	0.0099	0.0029	0.0026	0.0030
	FPA	0.0412	0.0312	0.0441	0.0148	0.0106	0.0235
	JA	0.0119	0.0089	0.0156	0.0086	0.0033	0.0166
	SSA	0.0452	0.0322	0.0508	0.0305	0.0239	0.0382
	DE	0.0311	0.0295	0.0352	0.0092	0.0050	0.0109
	COA	0.0360	0.0319	0.0372	0.0216	0.0172	0.0235
Heart	FA	0.0224	0.0182	0.0239	0.0105	0.0093	0.0113
	PSO	0.0290	0.0282	0.0299	0.0073	0.0033	0.0229
	GASI	0.0909	0.0909	0.0909	0.0932	0.0932	0.0932
	FPA	0.1625	0.1611	0.1635	0.1137	0.1031	0.1268
	JA	0.1026	0.0940	0.1252	0.1046	0.1023	0.1245
	SSA	0.1130	0.1008	0.1268	0.1258	0.1023	0.1512
	DE	0.1142	0.1054	0.1283	0.1106	0.1008	0.1268
Arrhythmia	COA	0.1018	0.1016	0.1023	0.1117	0.1039	0.1268
	FA	0.0920	0.0894	0.1023	0.1245	0.1245	0.1245
	PSO	0.1515	0.1512	0.1528	0.1107	0.1031	0.1268
	GASI	0.2157	0.2157	0.2157	0.3095	0.3079	0.3102
	FPA	0.3147	0.3030	0.3179	0.3409	0.2960	0.3684
	JA	0.3045	0.3019	0.3100	0.3527	0.3514	0.3589
	SSA	0.3156	0.3034	0.3183	0.3617	0.3608	0.3680
Sonar	DE	0.3009	0.2950	0.3036	0.3448	0.3020	0.3588
	COA	0.3072	0.3027	0.3170	0.3304	0.3170	0.3386
	FA	0.3106	0.3104	0.3109	0.3227	0.3106	0.3252
	PSO	0.3029	0.2946	0.3167	0.3377	0.3154	0.3597
	GASI	0.0624	0.0498	0.0655	0.0204	0.0204	0.0204
	FPA	0.0975	0.0828	0.1145	0.0725	0.0515	0.0845
	JA	0.0739	0.0508	0.0821	0.0673	0.0520	0.0824
Ionosphere	SSA	0.0766	0.0667	0.0986	0.0880	0.0674	0.1149
	DE	0.1070	0.0978	0.1133	0.0375	0.0201	0.0523
	COA	0.1205	0.1130	0.1301	0.0505	0.0356	0.0677
	FA	0.1081	0.0988	0.1142	0.0572	0.0371	0.0829
	PSO	0.0792	0.0501	0.0985	0.0592	0.0507	0.0681
	GASI	0.0405	0.0405	0.0405	0.0488	0.0423	0.0505
	FPA	0.0728	0.0607	0.0800	0.0732	0.0689	0.0800
Waveform	JA	0.0659	0.0595	0.0767	0.0548	0.0414	0.0598
	SSA	0.0999	0.0890	0.1077	0.0500	0.0405	0.0604
	DE	0.0874	0.0677	0.1062	0.0539	0.0405	0.0595
	COA	0.0666	0.0598	0.0785	0.0619	0.0508	0.0697
	FA	0.0653	0.0589	0.0700	0.0645	0.0589	0.0703
	PSO	0.0701	0.0683	0.0773	0.0559	0.0408	0.0607
	GASI	0.1196	0.1196	0.1196	0.1500	0.1459	0.1506
Waveform	FPA	0.1452	0.1405	0.1416	0.1618	0.1559	0.1688
	JA	0.1315	0.1308	0.1330	0.1537	0.1510	0.1579
	SSA	0.1438	0.1363	0.1512	0.1616	0.1511	0.1751
	DE	0.1352	0.1351	0.1363	0.1559	0.1543	0.1605
	COA	0.1338	0.1323	0.1361	0.1583	0.1554	0.1615
	FA	0.1416	0.1405	0.1462	0.1548	0.1528	0.1570
	PSO	0.1360	0.1333	0.1382	0.1511	0.1445	0.1597

4.6. Discussion

In this section, we delve into the main arguments that robustly demonstrate the performance of the suggested approach. These critical points not only provide a comprehensive understanding of the approach's effectiveness but also underscore its potential benefits. Through a detailed exploration of these arguments, we aim to establish the approach as a compelling and viable solution within its intended domain, offering valuable insights and outcomes for further consideration.

- A machine learning task requires an efficient feature selection approach that can choose the most optimal number of features and obtain a better performance. Using a wide range of features increases the

probability of selecting irrelevant and redundant attributes, which negatively influences the model's performance, while the strong reduction of the number of features risks losing the original information of the dataset. In this paper the proposed multi-objective fitness function allows both reducing the number of features and minimizing the classification error. Subsequently, the features selected from the cancer dataset present the maximum information for diagnostic or predictive tasks.

- The main goal of the suggested method is to take advantage of the best solutions obtained by many different SI-based feature selection approaches. This approach uses a genetic algorithm with a different strategy to develop a powerful feature selection technique that finds the best subset of features in many data sets from different fields. This strategy is based on an initial intelligent population composed of the best solutions obtained by the different SI-based feature selection approaches. In addition, the genetic operators (crossover mutation selection) keep the diversity of the generation to enhance the quality of the search space exploration and avoid the local optimum problem.
- The temporal complexity is not a real obstacle because the selection of the characteristics is made with the exploitation of the model. This preliminary stage will not be repeated with each use of the machine learning model.

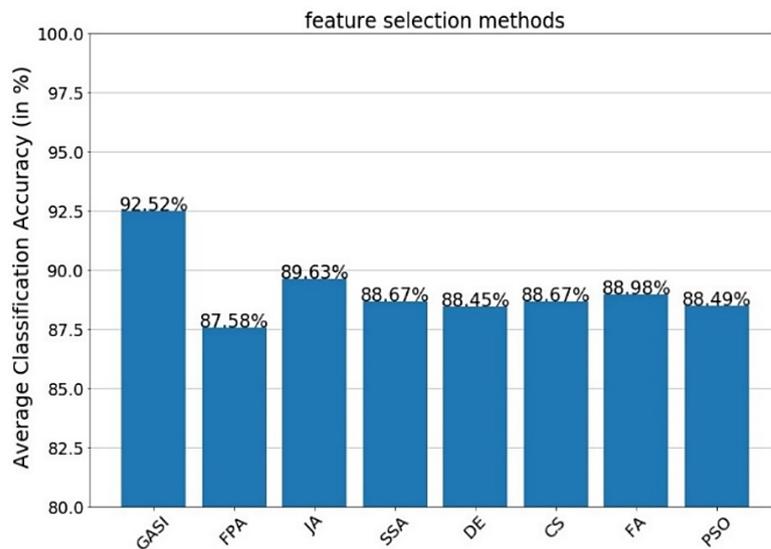


Figure 3. Average classification accuracy over all datasets on the Logistic regression classifier

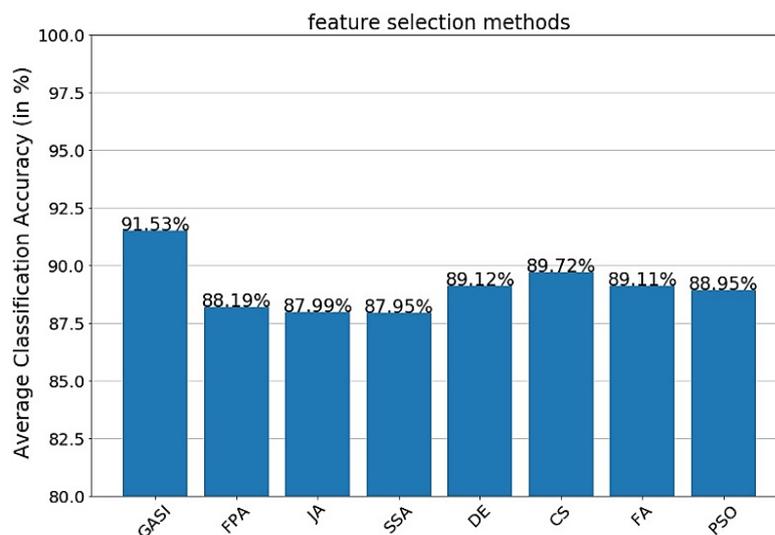


Figure 4. Average classification accuracy over all datasets on the AdaBoost classifier

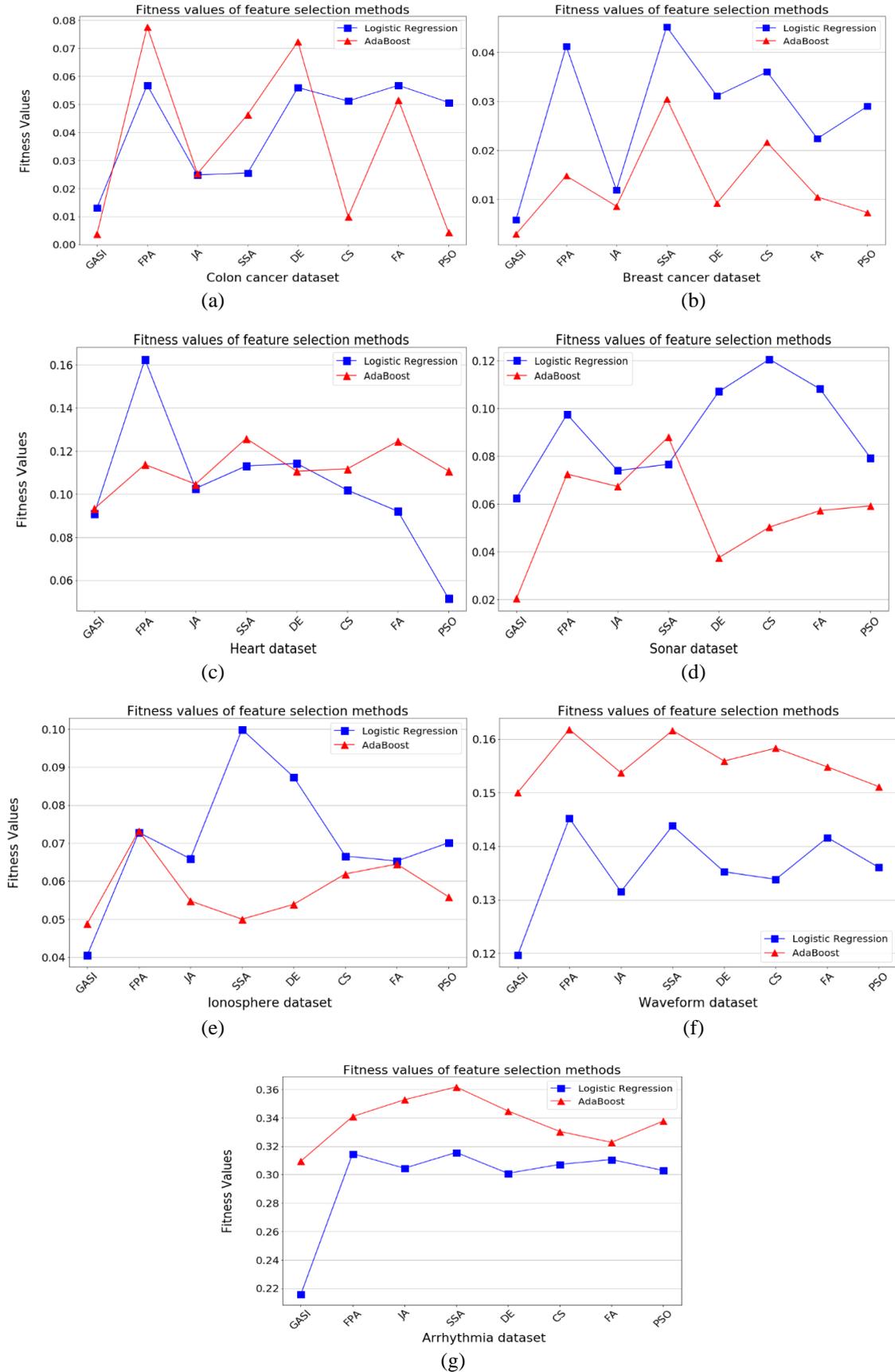


Figure 5. Average fitness values of feature selection methods across diverse datasets (a) colon cancer, (b) breast cancer, (c) heart, (d) sonar, (e) ionosphere, (f) waveform, and (g) arrhythmia

5. CONCLUSION

With the massive amounts of digital data of various types and the exponential growth of artificial intelligence-based applications, the size of the data is increasing, leading to massive databases with a large number of features, especially in the medical field. At the same time, data mining and machine learning tasks require fast speed and greater accuracy. Over the past few years, numerous meta-heuristic methods have been developed to reduce the size of the dataset by eliminating redundant and irrelevant features that represent noise for the model. This paper suggests a novel powerful feature selection method, which uses a strategy that combines many SI-based (i.e., FPA, JA, SSA, DE, COA, FA, and PSO) feature selection approaches and employs a genetic algorithm that uses a multi-objective fitness function to discover the optimal subset of features in many data sets from different areas. This approach is applied to seven well-known datasets from the UCI repository for feature selection. The results obtained were compared with many powerful different SI-based feature selection approaches, and the experiments show that our method obtained better solutions in terms of fitness value and classification accuracy. Day by day, world health is affected by numerous invasive pathologies, especially heart disorders and cancer. This study shows the necessity of raising healthcare professionals' awareness about the efficient use of powerful feature selection techniques that may be successfully applied to medical databases for detecting, classifying, and predicting diseases. For future work, the suggested technique can be employed in high-dimensional datasets, and it can be combined with other metaheuristic techniques to more effectively improve the exploration of the searching space and accelerate convergence. Moreover, the suggested approach can also be used to solve various real-world problems.

REFERENCES

- [1] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: 10.1016/j.neucom.2017.11.077.
- [2] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, Sep. 2019, doi: 10.1016/j.combiomed.2019.103375.
- [3] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter-wrapper feature selection for short-term load forecasting," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, Apr. 2015, doi: 10.1016/j.engappai.2014.12.014.
- [4] R. Guha *et al.*, "Deluge based genetic algorithm for feature selection," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 357–367, Jun. 2021, doi: 10.1007/s12065-019-00218-5.
- [5] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems," *Engineering Applications of Artificial Intelligence*, vol. 70, pp. 25–37, Apr. 2018, doi: 10.1016/j.engappai.2017.12.014.
- [6] J. Zhang, Y. Xiong, and S. Min, "A new hybrid filter/wrapper algorithm for feature selection in classification," *Analytica Chimica Acta*, vol. 1080, pp. 43–54, Nov. 2019, doi: 10.1016/j.aca.2019.06.054.
- [7] N. El Aboudi and L. Benhlila, "Review on wrapper feature selection approaches," in *2016 International Conference on Engineering and MIS (ICEMIS)*, Sep. 2016, pp. 1–5, doi: 10.1109/ICEMIS.2016.7745366.
- [8] Y. Fu, X. Liu, S. Sarkar, and T. Wu, "Gaussian mixture model with feature selection: An embedded approach," *Computers & Industrial Engineering*, vol. 152, p. 107000, Feb. 2021, doi: 10.1016/j.cie.2020.107000.
- [9] H. Faris *et al.*, "Time-varying hierarchical chains of salps with random weight networks for feature selection," *Expert Systems with Applications*, vol. 140, Feb. 2020, doi: 10.1016/j.eswa.2019.112898.
- [10] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzandeh, "Review of swarm intelligence-based feature selection methods," *Engineering Applications of Artificial Intelligence*, vol. 100, Apr. 2021, doi: 10.1016/j.engappai.2021.104210.
- [11] A. E. Hegazy, M. A. Makhlof, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 3, pp. 335–344, Mar. 2020, doi: 10.1016/j.jksuci.2018.06.003.
- [12] S. Forouzandeh, K. Berahmand, and M. Rostami, "Presentation of a recommender system with ensemble learning and graph embedding: a case on MovieLens," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7805–7832, Feb. 2021, doi: 10.1007/s11042-020-09949-5.
- [13] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the Gini index and information gain criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, no. 1, pp. 77–93, May 2004, doi: 10.1023/B:AMAI.0000018580.96245.c6.
- [14] J.-H. Cheng, D.-W. Sun, and H. Pu, "Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen-thawed fish muscle," *Food Chemistry*, vol. 197, pp. 855–863, Apr. 2016, doi: 10.1016/j.foodchem.2015.11.019.
- [15] P. Hanchuan, L. Fuhui, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.
- [16] A. J. Ferreira and M. A. T. Figueiredo, "An unsupervised approach to feature discretization and selection," *Pattern Recognition*, vol. 45, no. 9, pp. 3048–3060, Sep. 2012, doi: 10.1016/j.patcog.2011.12.008.
- [17] N. D. Cilia, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "Variable-length representation for EC-based feature selection in high-dimensional data," in *Applications of Evolutionary Computation*, Springer International Publishing, 2019, pp. 325–340.
- [18] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, doi: 10.1016/S0004-3702(97)00043-X.
- [19] T. N. Lal, O. Chapelle, J. Weston, and A. Elisseeff, "Embedded methods," in *Feature Extraction*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 137–165.
- [20] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112–123, Jun. 2014, doi: 10.1016/j.engappai.2014.03.007.
- [21] S. Barak, J. H. Dahooie, and T. Tichý, "Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9221–9235, Dec. 2015, doi: 10.1016/j.eswa.2015.08.010.
- [22] C. Wang, H. Pan, and Y. Su, "A many-objective evolutionary algorithm with diversity-first based environmental selection,"

- Swarm and Evolutionary Computation*, vol. 53, Mar. 2020, doi: 10.1016/j.swevo.2019.100641.
- [23] A. AlSukkk, R. N. Khushaba, and A. Al-Ani, "Enhancing the diversity of genetic algorithm for improved feature selection," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, Oct. 2010, pp. 1325–1331, doi: 10.1109/ICSMC.2010.5642445.
- [24] F. Amini and G. Hu, "A two-layer feature selection method using genetic algorithm and Elastic Net," *Expert Systems with Applications*, vol. 166, Mar. 2021, doi: 10.1016/j.eswa.2020.114072.
- [25] F. Tan, X. Fu, Y. Zhang, and A. G. Bourgeois, "A genetic algorithm-based method for feature subset selection," *Soft Computing*, vol. 12, no. 2, pp. 111–120, Sep. 2007, doi: 10.1007/s00500-007-0193-8.
- [26] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 87–103, Mar. 2016, doi: 10.1016/j.ymssp.2015.08.030.
- [27] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.
- [28] P. Dhal and C. Azad, "A multi-objective feature selection method using Newton's law based PSO with GWO," *Applied Soft Computing*, vol. 107, Aug. 2021, doi: 10.1016/j.asoc.2021.107394.
- [29] H. T. Ibrahim, W. J. Mazher, and E. M. Jassim, "Modified Harris Hawks optimizer for feature selection and support vector machine kernels," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 942–953, Feb. 2023, doi: 10.11591/ijeecs.v29.i2.pp942-953.
- [30] S. Masrom, R. A. Rahman, M. Mohamad, A. S. Abd Rahman, and N. Baharun, "Machine learning of tax avoidance detection based on hybrid metaheuristics algorithms," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 3, pp. 1153–1163, Sep. 2022, doi: 10.11591/ijai.v11.i3.pp1153-1163.
- [31] A. N. Jaber, K. Moorthy, L. Machap, and S. Deris, "The importance of data classification using machine learning methods in microarray data," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 491–498, Apr. 2021, doi: 10.12928/telkomnika.v19i2.15948.
- [32] K. Gangadharan, G. R. N. Kumari, D. Dhanasekaran, and K. Malathi, "Detection and classification of various pest attacks and infection on plants using recursive backpropagation neural network with GA based particle swarm optimization algorithm," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 20, no. 3, pp. 1278–1288, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1278-1288.
- [33] Y. Xue, T. Tang, W. Pang, and A. X. Liu, "Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifiers," *Applied Soft Computing*, vol. 88, Mar. 2020, doi: 10.1016/j.asoc.2019.106031.
- [34] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [35] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, and J. Xu, "An improved differential evolution algorithm and its application in optimization problem," *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, Apr. 2021, doi: 10.1007/s00500-020-05527-x.
- [36] Y. Zhang, D. Gong, X. Gao, T. Tian, and X. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Information Sciences*, vol. 507, pp. 67–85, Jan. 2020, doi: 10.1016/j.ins.2019.08.040.
- [37] H. Li, F. He, Y. Chen, and Y. Pan, "MLFS-CCDE: multi-objective large-scale feature selection by cooperative coevolutionary differential evolution," *Memetic Computing*, vol. 13, no. 1, pp. 1–18, Mar. 2021, doi: 10.1007/s12293-021-00328-7.
- [38] R. Rajabioun, "Cuckoo optimization algorithm," *Applied Soft Computing*, vol. 11, no. 8, pp. 5508–5518, Dec. 2011, doi: 10.1016/j.asoc.2011.05.008.
- [39] M. Kelidari and J. Hamidzadeh, "Feature selection by using chaotic cuckoo optimization algorithm with levy flight, opposition-based learning and disruption operator," *Soft Computing*, vol. 25, no. 4, pp. 2911–2933, Feb. 2021, doi: 10.1007/s00500-020-05349-x.
- [40] V. Jayaraman and H. P. Sultana, "Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification," *Journal of Ambient Intelligence and Humanized Computing*, Jan. 2019, doi: 10.1007/s12652-019-01193-6.
- [41] X.-S. Yang, "Firefly algorithm, Lévy flights and global optimization," in *Research and Development in Intelligent Systems XXVI*, London: Springer London, 2010, pp. 209–218.
- [42] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163–191, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.
- [43] G. I. Sayed, G. Khoriba, and M. H. Haggag, "A novel chaotic salp swarm algorithm for global optimization and feature selection," *Applied Intelligence*, vol. 48, no. 10, pp. 3462–3481, Oct. 2018, doi: 10.1007/s10489-018-1158-6.
- [44] M. Tubishat *et al.*, "Dynamic salp swarm algorithm for feature selection," *Expert Systems with Applications*, vol. 164, Feb. 2021, doi: 10.1016/j.eswa.2020.113873.
- [45] A. Chaudhuri and T. P. Sahu, "A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification," *Computers and Electrical Engineering*, vol. 90, Mar. 2021, doi: 10.1016/j.compeleceng.2020.106963.
- [46] X.-S. Yang, "Flower pollination algorithm for global optimization," in *Unconventional Computation and Natural Computation*, Springer Berlin Heidelberg, 2012, pp. 240–249.
- [47] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016, doi: 10.1016/j.neucom.2015.06.083.
- [48] P. Agrawal, T. Ganesh, and A. W. Mohamed, "Chaotic gaining sharing knowledge-based optimization algorithm: an improved metaheuristic algorithm for feature selection," *Soft Computing*, vol. 25, no. 14, pp. 9505–9528, Jul. 2021, doi: 10.1007/s00500-021-05874-3.
- [49] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 6, pp. 2193–2196, Mar. 2012, doi: 10.1016/j.physa.2011.12.004.
- [50] B. Yeganeh *et al.*, "Targeting the mevalonate cascade as a new therapeutic approach in heart disease, cancer and pulmonary disease," *Pharmacology & Therapeutics*, vol. 143, no. 1, pp. 87–110, Jul. 2014, doi: 10.1016/j.pharmthera.2014.02.007.
- [51] W. Książek, M. Gandor, and P. Plawiak, "Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma," *Computers in Biology and Medicine*, vol. 134, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104431.
- [52] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

- [53] Y. Liang, S. Zhang, H. Qiao, and Y. Yao, "iPromoter-ET: Identifying promoters and their strength by extremely randomized trees-based feature selection," *Analytical Biochemistry*, vol. 630, Oct. 2021, doi: 10.1016/j.ab.2021.114335.
- [54] C. Peng, X. Wu, W. Yuan, X. Zhang, Y. Zhang, and Y. Li, "MGRFE: multilayer recursive feature elimination based on an embedded genetic algorithm for cancer classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 2, pp. 621–632, Mar. 2021, doi: 10.1109/TCBB.2019.2921961.

BIOGRAPHIES OF AUTHORS



Salsabila Benghazouani    obtained a master's degree in computer engineering from the Faculty of Science and Technology, Mohammedia, Morocco, in 2006, and a master's degree in data science from the Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, Morocco, in 2021. She is currently preparing her Ph.D. in computer science at the TIM Lab, Ben M'Sik Faculty of Science, Hassan II University, Casablanca, Morocco. Her research focuses on artificial intelligence, machine learning and deep learning. She can be contacted at email: benghazouani.salsabila239@gmail.com.



Said Nough    holds a Ph.D. in computer sciences at National School of Computer Science and Systems Analysis (ENSIAS), Rabat, Morocco in 2014. He is currently professor (Higher Degree Research (HDR)) at Faculty of Sciences Ben M'Sick, Hassan II University, Casablanca, Morocco. His current research interests are artificial intelligence, machine learning, deep learning, telecommunications, information and coding theory. He can be contacted at email: said.nough@univh2m.ma.



Abdelali Zakrani    obtained his B.Sc. and DESA (M.Sc.) degrees in computer science from Hassan II University, Casablanca, Morocco, in 2003 and 2005, respectively, and the Ph.D. degree in the same field from Mohammed V University, Rabat, Morocco, in 2012. His current research interests' artificial neural network, data mining and software engineering. He can be contacted at email: abdelali.zakrani@univh2c.ma.



Ihsane Haloum    is currently a Ph.D. student, affiliated with the Laboratory of Immuno-genetics and Human Pathologies, at the Faculty of Medicine and Pharmacy of Casablanca, Hassan II University of Casablanca, Morocco. She received her M.Sc. degree in medical biotechnologies, in 2021, from the Faculty of Medicine and Pharmacy of Rabat, Morocco. Her research topics are related to immunogenetics, oncology, and artificial intelligence. She can be contacted at email: ihsane.haloum@gmail.com.



Mostafa Jebbar    holds a Ph.D. in computer science. He is currently associate professor at Superior School of Technology, Hassan II University, Casablanca, Morocco. His current research interests are architecture logicielle, RM-ODPSOA, cloud computing. He can be contacted at email: mostafajebbar@gmail.com.